DISCLAIMER
Reference herein to any specific commercial company, product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the Department of the Army (DoA). The opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or the DoA, and shall not be used for advertising or product endorsement purposes.

DISTRIBUTION STATEMENT A. Approved for public release, distribution is unlimited. OPSEC # 23238

COMBAT ID – Combat Identification System

Garbis Salgian, Zsolt Kira, Raia Hadsell, Han-Pang Chiu, Xun Zhou, Bing-Bing Chai, Supun Samarasekera
SRI International, Princeton, NJ

Bernard Theisen
Jeffery Ramsey
U.S. Army TARDEC, Warren, MI

9 August 2012

GVSETS

Report Documentation Page

Form Approved
OMB No. 0704-0188

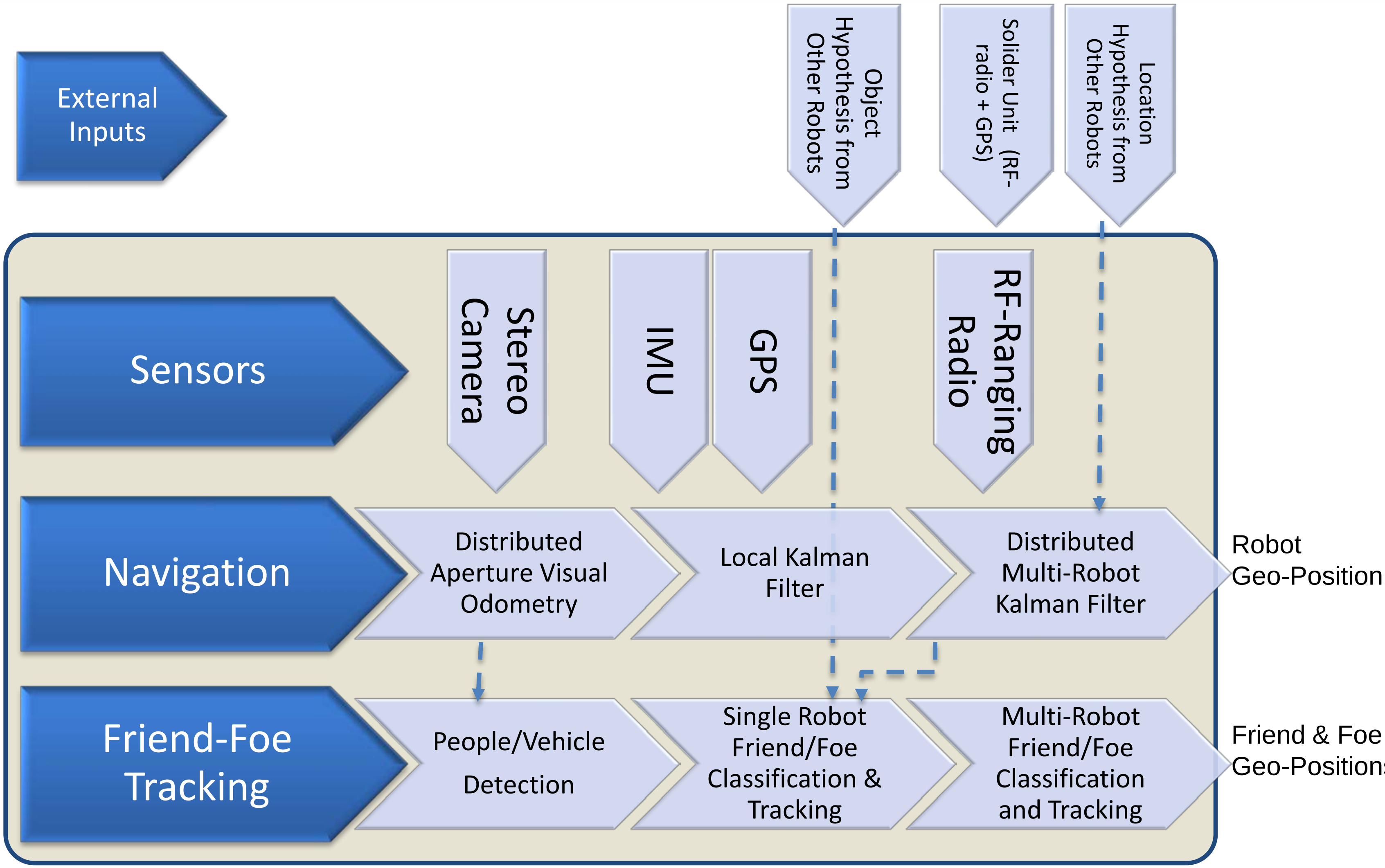
1. REPORT DATE	**02 AUG 2012**
2. REPORT TYPE	**Briefing**
3. DATES COVERED	**01-07-2012 to 01-08-2012**
4. TITLE AND SUBTITLE	**COMBAT ID Combat Identification System**
5a. CONTRACT NUMBER	
5b. GRANT NUMBER	
5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)	**Bernard Theisen; Jeffery Ramsey; Xun Zhou; Supun Samarasekera; Zsolt Kira**
5d. PROJECT NUMBER	
5e. TASK NUMBER	
5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**U.S. Army TARDEC,6501 East Eleven Mile Rd,Warren,Mi,48397-5000**
8. PERFORMING ORGANIZATION REPORT NUMBER	**#23238**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**U.S. Army TARDEC, 6501 East Eleven Mile Rd, Warren, Mi, 48397-5000**
10. SPONSOR/MONITOR'S ACRONYM(S)	**TARDEC**
11. SPONSOR/MONITOR'S REPORT NUMBER(S)	**#23238**
12. DISTRIBUTION/AVAILABILITY STATEMENT	**Approved for public release; distribution unlimited**
13. SUPPLEMENTARY NOTES	**Submitted to 2012 NDIA Ground Vehicle Systems Engineering and Technology Symposium August 14-16 Troy, Michigan**
14. ABSTRACT	**-Integrate, test and demonstrate a fully integrated hardware and software solution running on two robot systems and three additional blue force entities. -Reliably detect blue and red force entities within a 60m radius, 180deg around each robot. -The proposed solution is designed to run through multiple classes of robot systems starting from Small UGV?s through large vehicles such as trucks or tanks.**
15. SUBJECT TERMS	

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Public Release**	**38**	

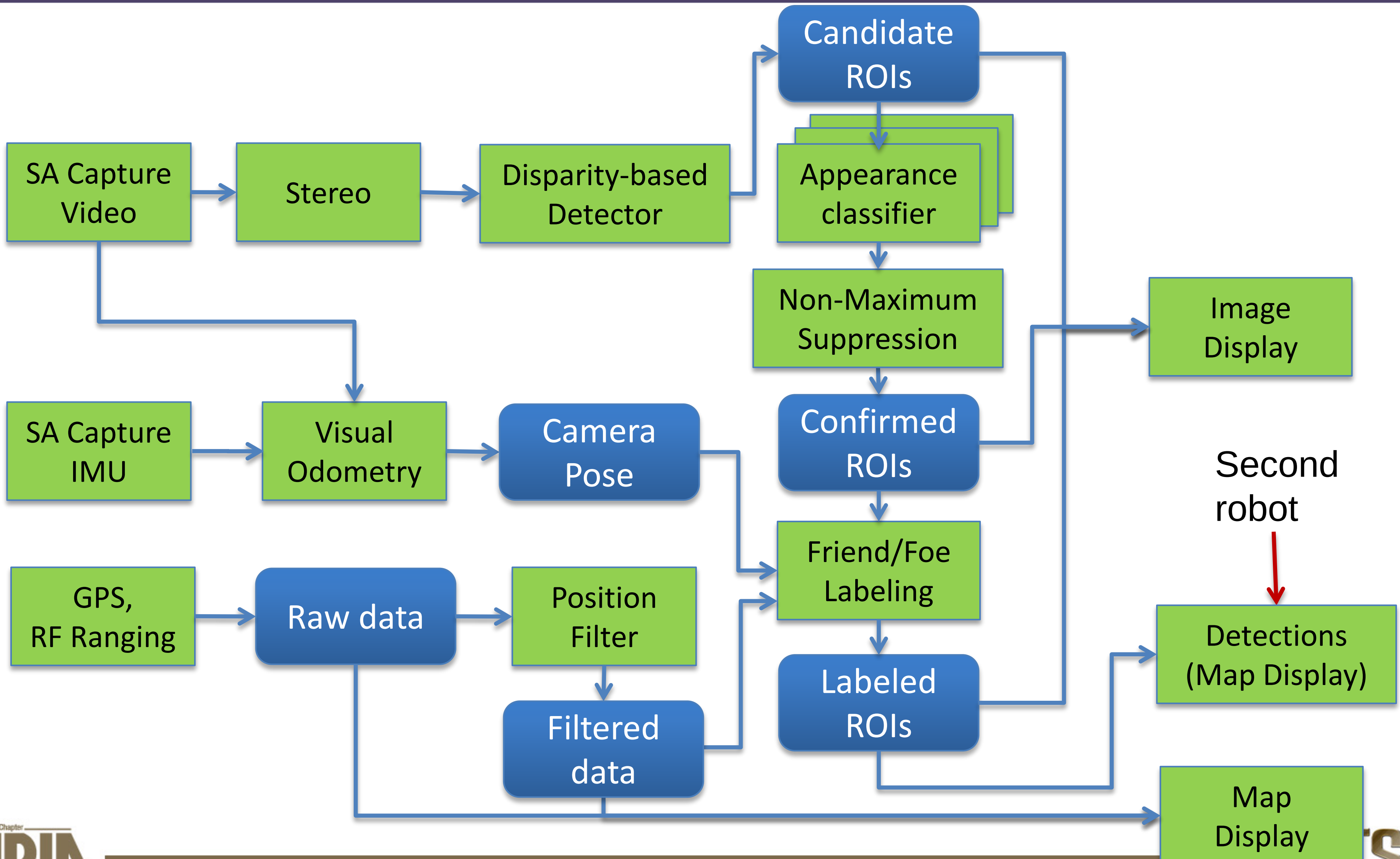
Project scope

- Integrate, test and demonstrate a fully integrated hardware and software solution running on two robot systems and three additional blue force entities.
- Reliably detect blue and red force entities within a 60m radius, 180deg around each robot.
- The proposed solution is designed to run through multiple classes of robot systems starting from Small UGV's through large vehicles such as trucks or tanks.

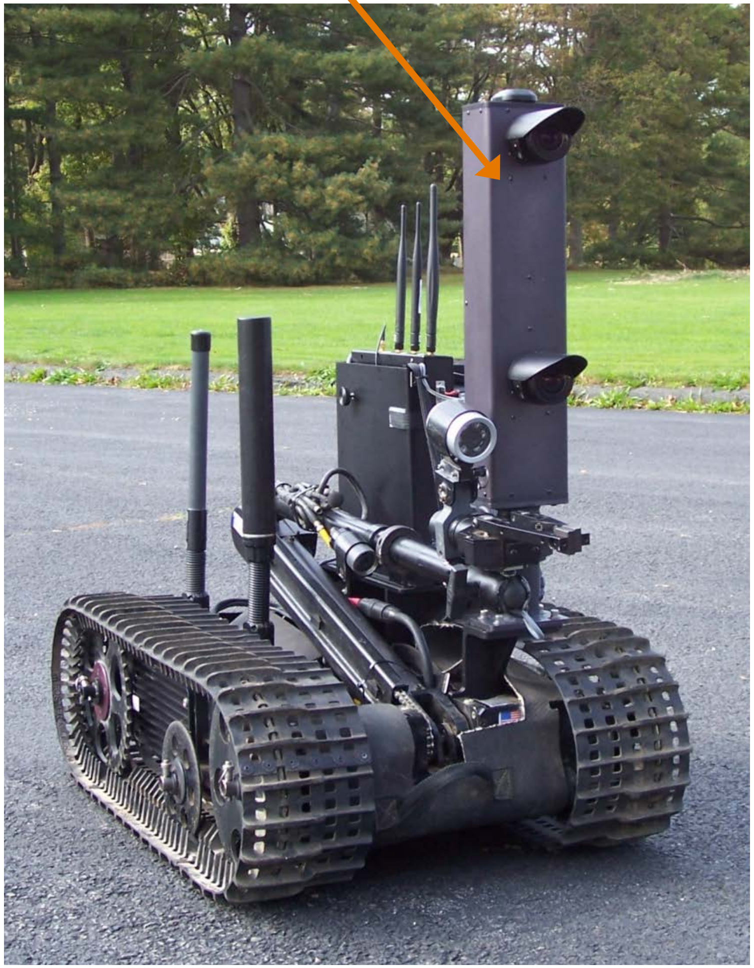
Overall System Design

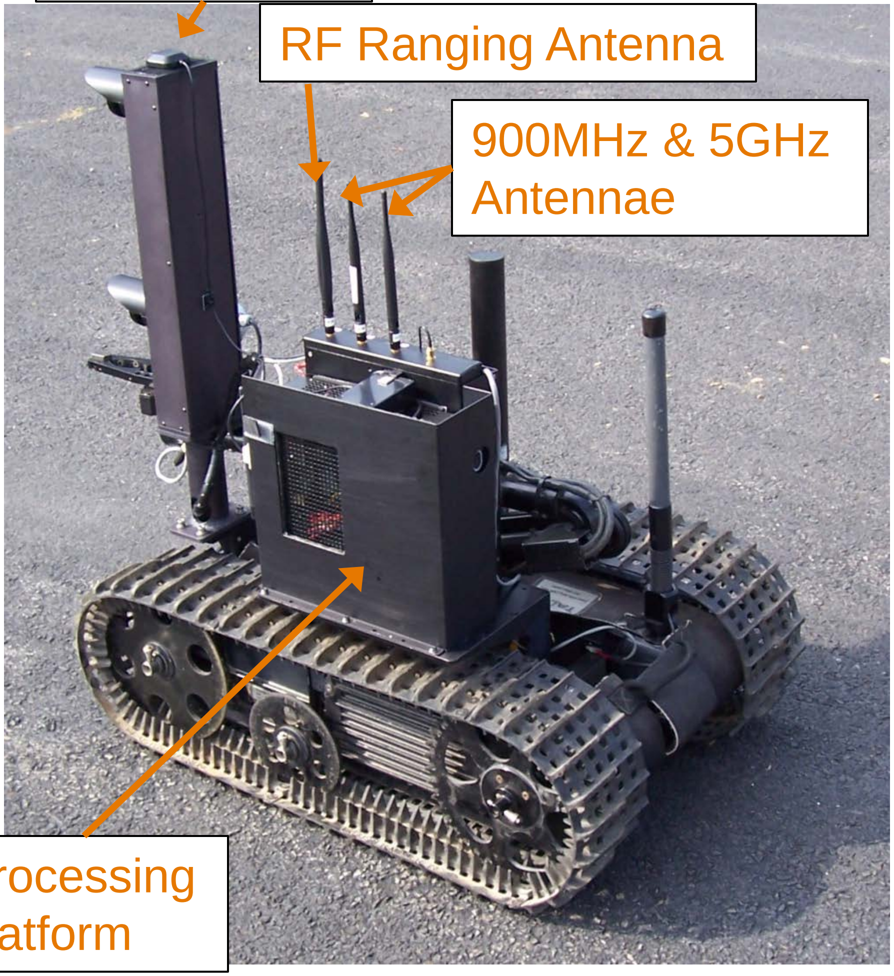


Block Diagram

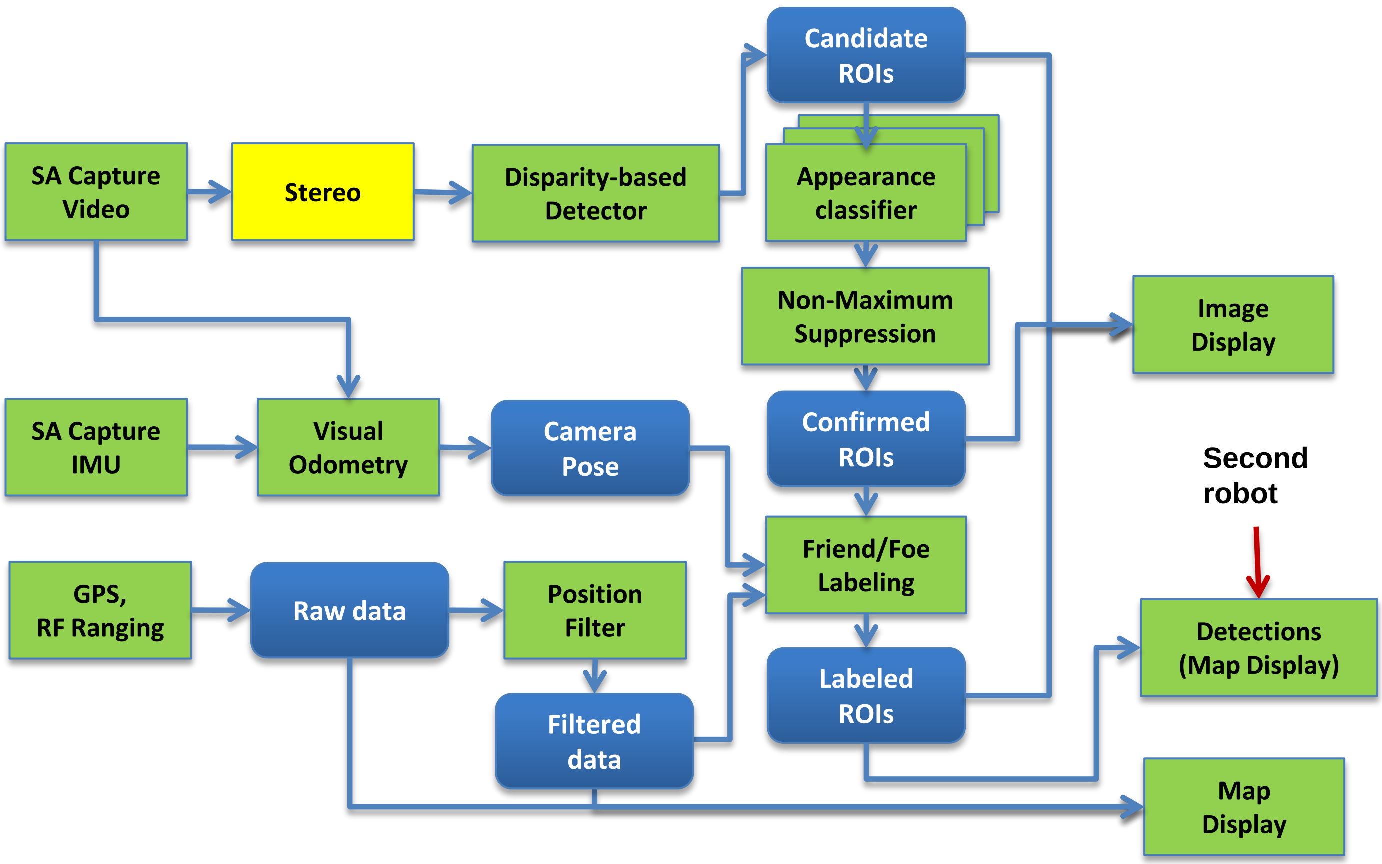


System installed on TALON

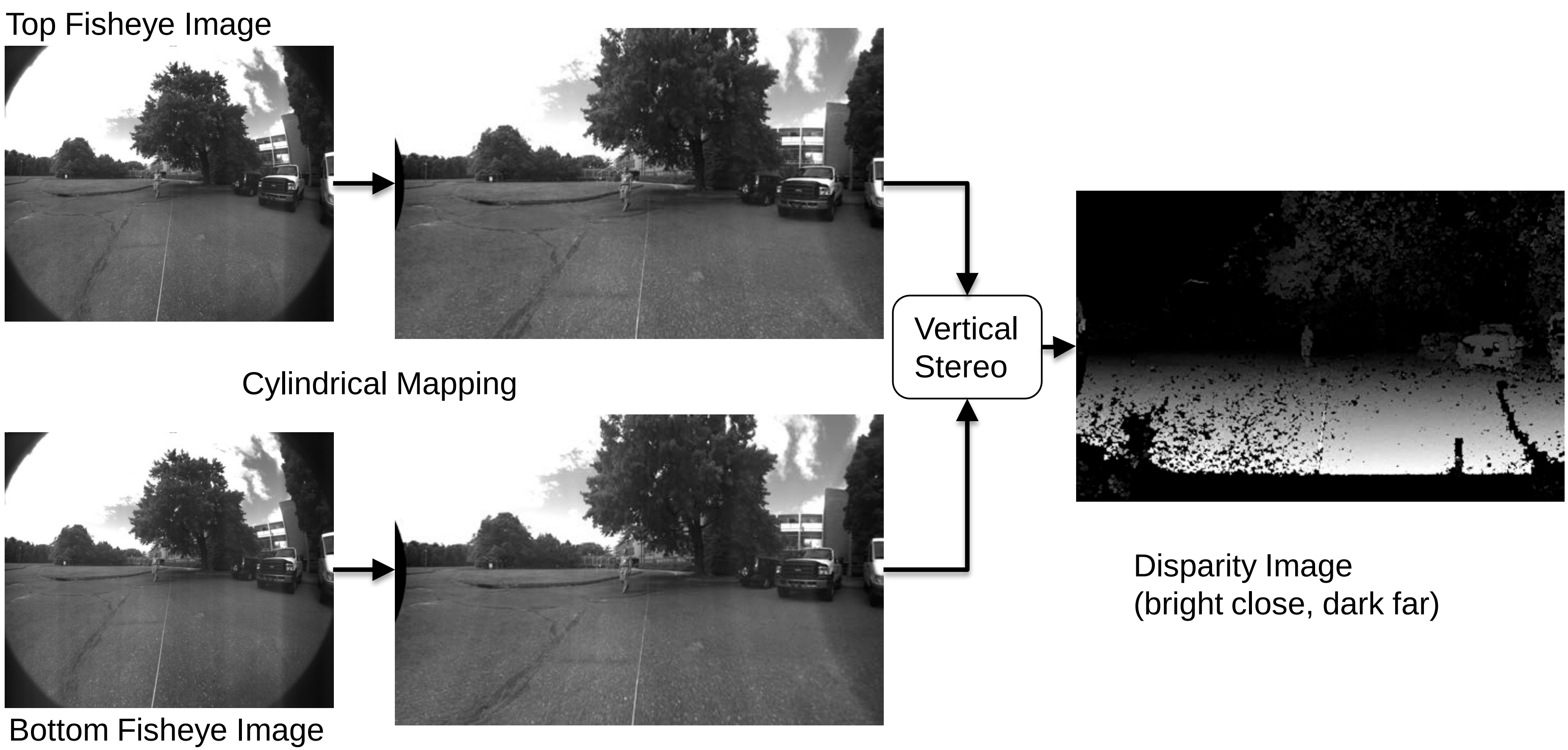




Stereo

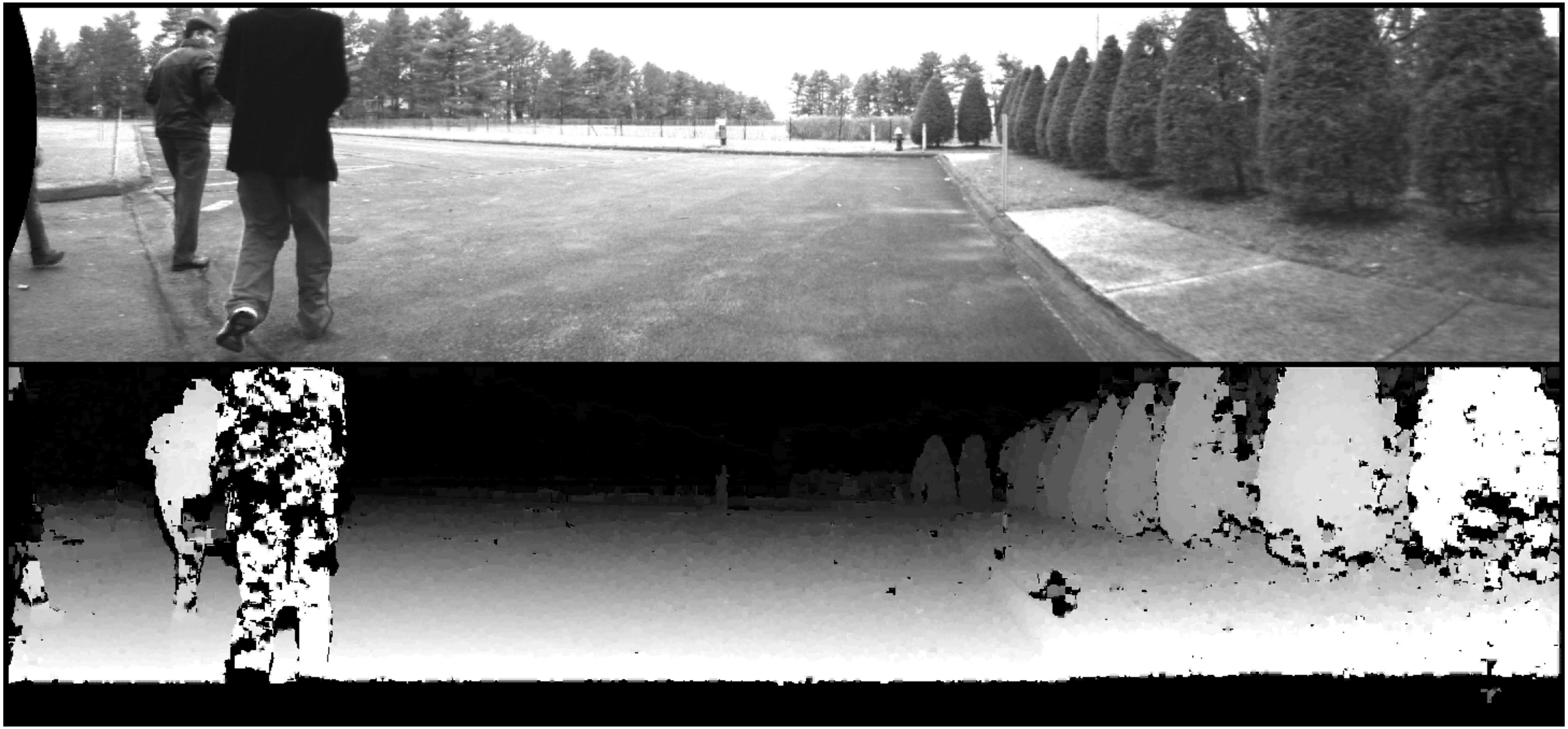

SA Capture Video
Stereo
Disparity-based Detector
Candidate ROIs
Appearance classifier
Non-Maximum Suppression
Confirmed ROIs
Friend/Foe Labeling
Labeled ROIs
Image Display
Second robot
Detections (Map Display)
Map Display
SA Capture IMU
Visual Odometry
Camera Pose
GPS, RF Ranging
Raw data
Position Filter
Filtered data

Fisheye Vertical Stereo



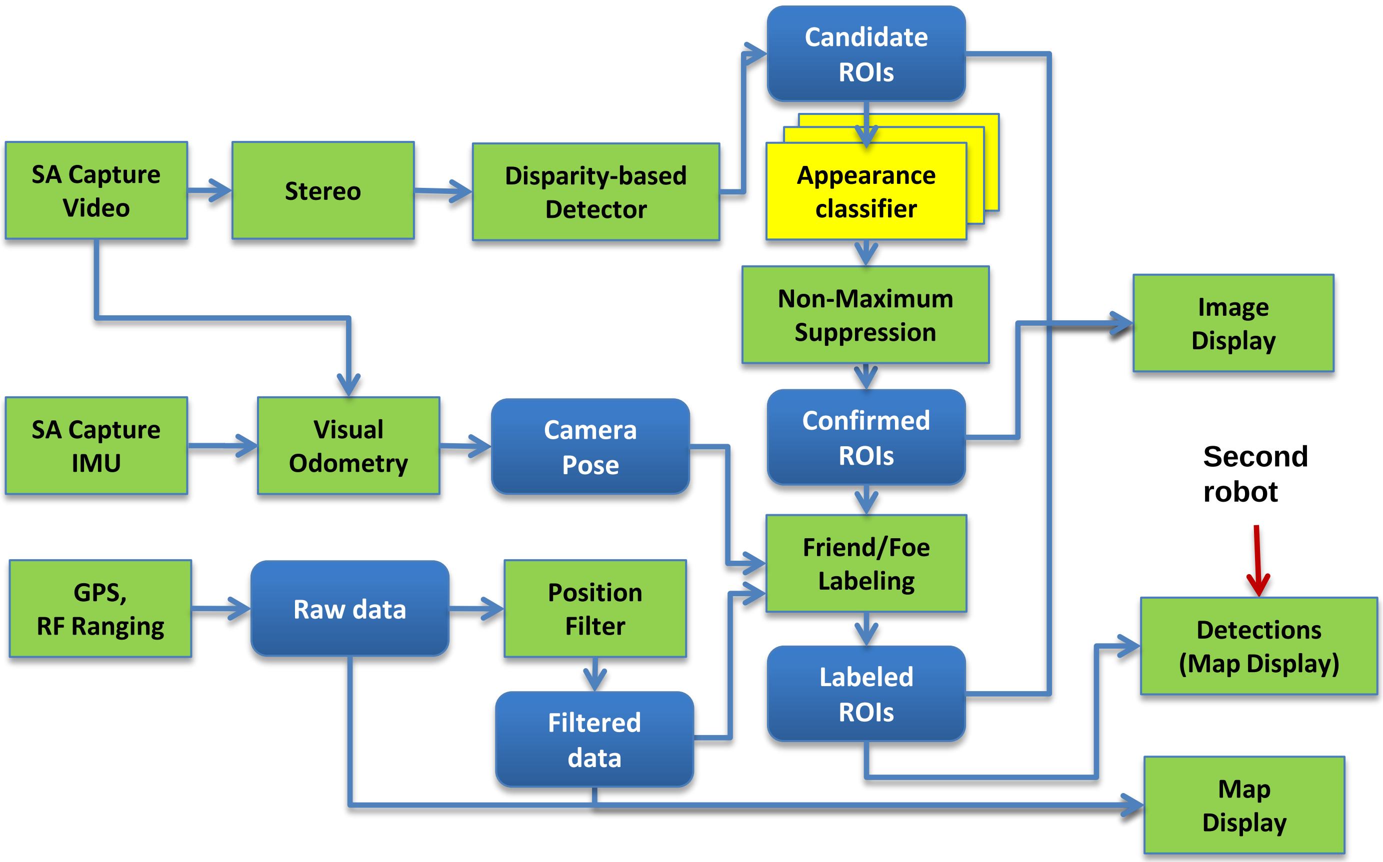
Fisheye vertical stereo example

Stereo Reference Image



Disparity Image (closer points are brighter)

Appearance Classifiers

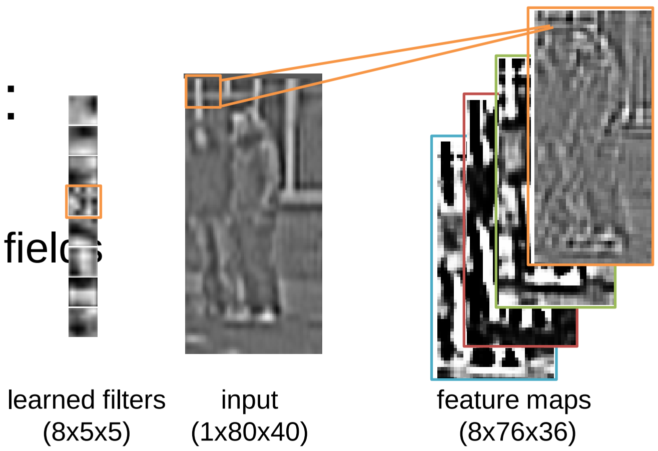


Person Classification

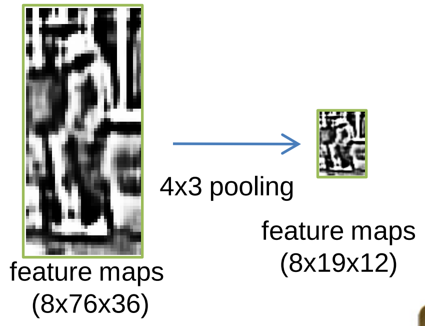
- **Accurate Person Recognition** is difficult because of low numbers of pixels on target, deformation and articulation, and shadows/glare.
- There are many modern approaches for person/pedestrian classification.
  - All of these use statistical learning methods to recognize patterns in the input.
  - However, none is perfect (less than 1 false positive per frame is "excellent" performance), because of the inherent difficulty of the task.
- We use **Hierarchical Feature Learning** to automatically learn custom features and a classifier directly from data.
- This is a fully supervised learning method, so it relies on a broad array of annotated ground truth data. We hand-labeled 25 video sequences for this purpose.
- The Learning architecture is called a Convolutional Neural Net, and is described on the next slide.

Hierarchical Feature Extraction

- Convolutional Neural Networks (ConvNets) are one method for simultaneous feature learning and classifier training. Since they involve training multiple, stacked non-linear transforms, they are considered an architecture for Deep Learning.
- ConvNet architectural components:
  - convolution layers
    - extract features using small local receptive fields
    - detect patterns with increasing complexity
    - use spatial or temporal weight-sharing
    - allow complex, nonlinear transformations



  - subsampling layers
    - pool features by local averaging
    - increase shift and scale invariance
    - reduce computational complexity

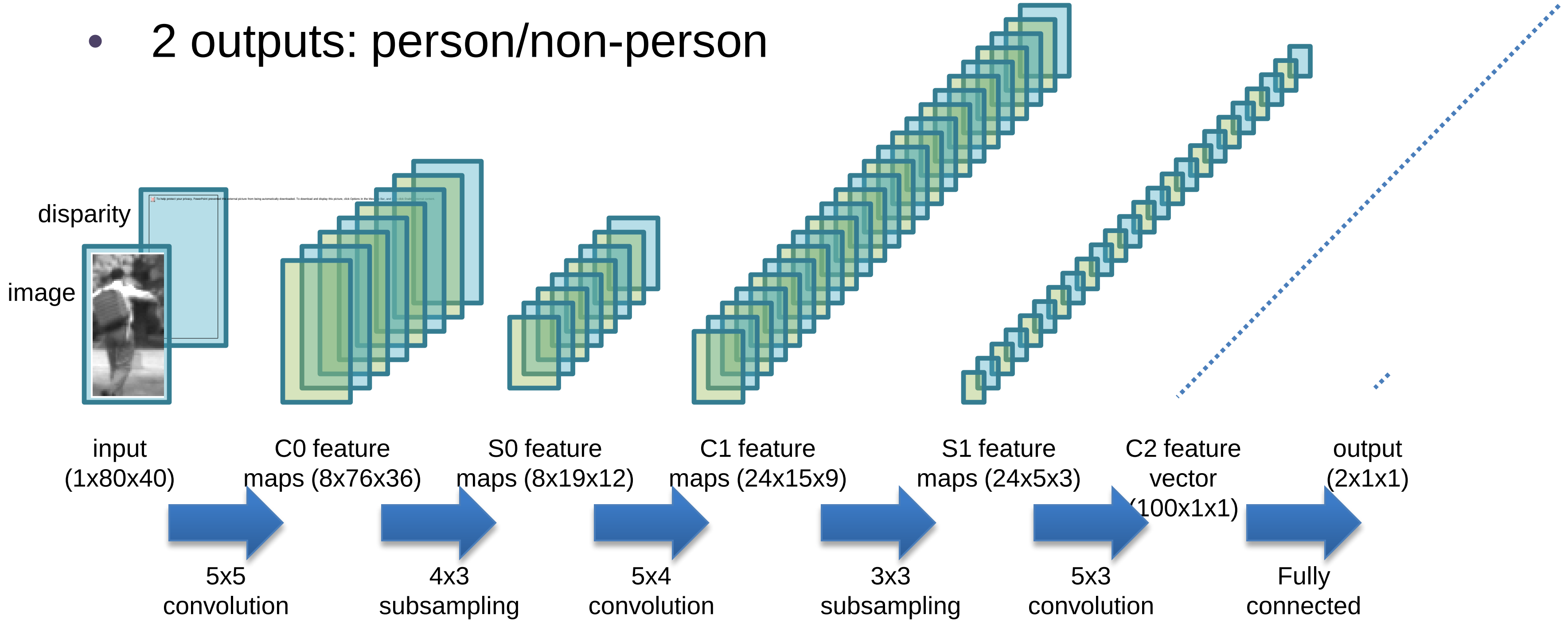


Person Classification

- **Our solution**: After comparison with other state-of-the-art methods, a Convolutional Neural Network (ConvNet) was chosen
  - Uses 2 inputs: appearance and disparity map
- **Network details**:
  - Modeled after similar architectures built for autonomous navigation (LAGR) and handwriting recogntion (LeNet5)
    - 6 layer hierarchy (3 convolutional layers, 2 pooling layers, and a fully connected layer)
    - 80x40 pixel field of view with dual input layers
      - 1st layer: normalized 8bit grayscale
      - 2nd layer: normalized disparities
    - 8,000 trainable parameters.
- **Training process**: Based on human-annotated videos
  - 800,000 labeled **positives** (ROIs with vehicles) and **negatives** (ROIs with no vehicles)
  - Network parameters are optimized using stochastic gradient descent

Convolutional Neural Network Architecture for Pedestrian Classifier

- 6 layer network with dual input layers: image and disparity
- 2 outputs: person/non-person



Pedestrian: Dataset examples of image input layer

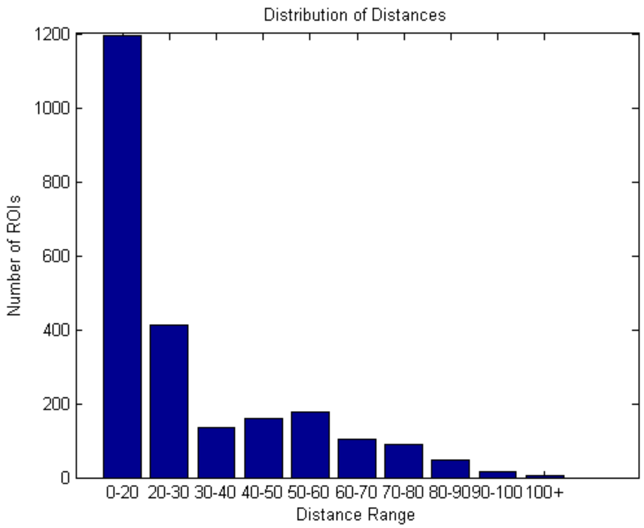
Pedestrian: Dataset examples of disparity input layer

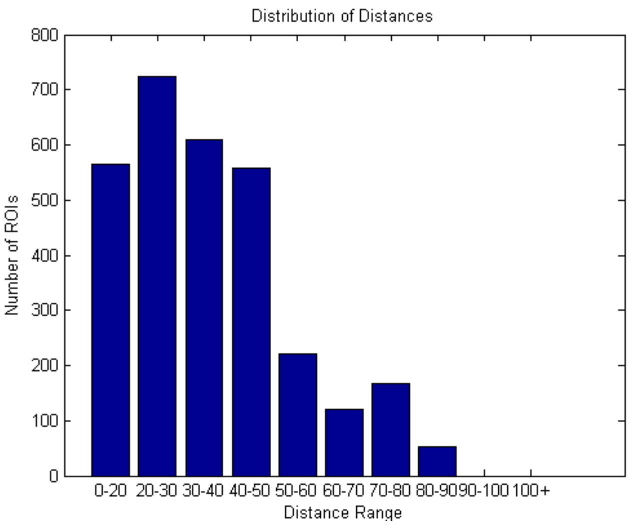
Pedestrian Classification Results

- We have performed extensive testing of the pedestrian classifier over datasets taken throughout the year
  - Each dataset contains 4-6 collections gathered in different environments including open areas, parking lot, and forest.
- Metrics - We used standard metrics used in the literature:
  - Recall is the ratio of positive detections and *all actual positives* in the dataset. This measures how well the classifier picks up people.
  - Precision is the ratio of true positives and *all detections* returned by the classifier. This measures how specific the classifier's detections are to people.
  - False positives per image (FPPI) is the mean number of false positives per image.

Pedestrian Classification Results

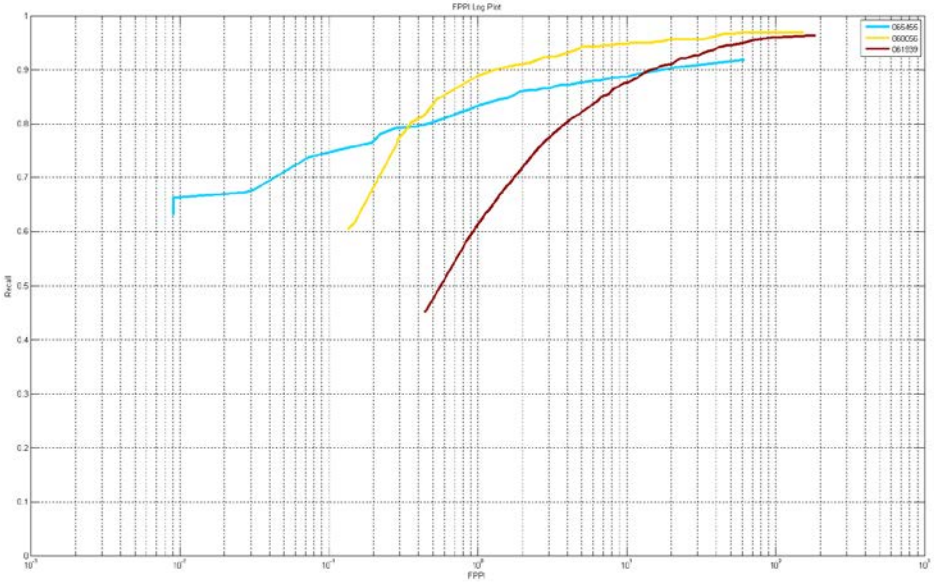
- We have performed extensive testing of the pedestrian classifier over datasets taken throughout the year
  - Each dataset contains 4-6 collections gathered in different environments including open areas, parking lot, and forest.
- Dataset: 2011.06.06: Fisheye and 80 degree
- Five sequences, both stationary and moving camera



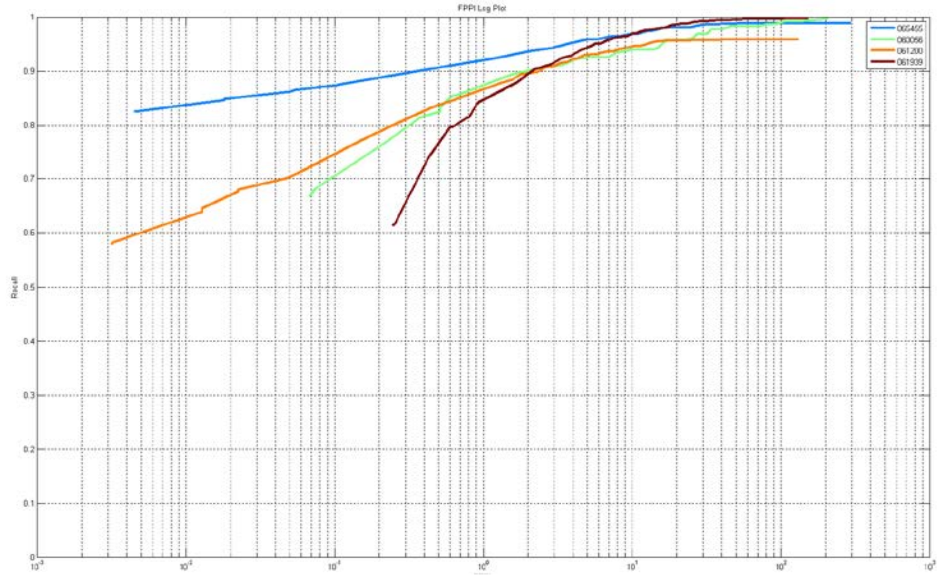


Pedestrian Classification Results

- Metrics - We used standard metrics used in the literature:
  - Recall is the ratio of positive detections and *all actual positives* in the dataset. This measures how well the classifier picks up people.
  - False positives per image (FPPI) is the mean number of false positives per image.

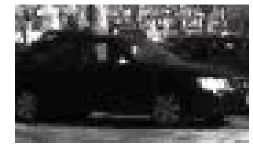


Fisheye



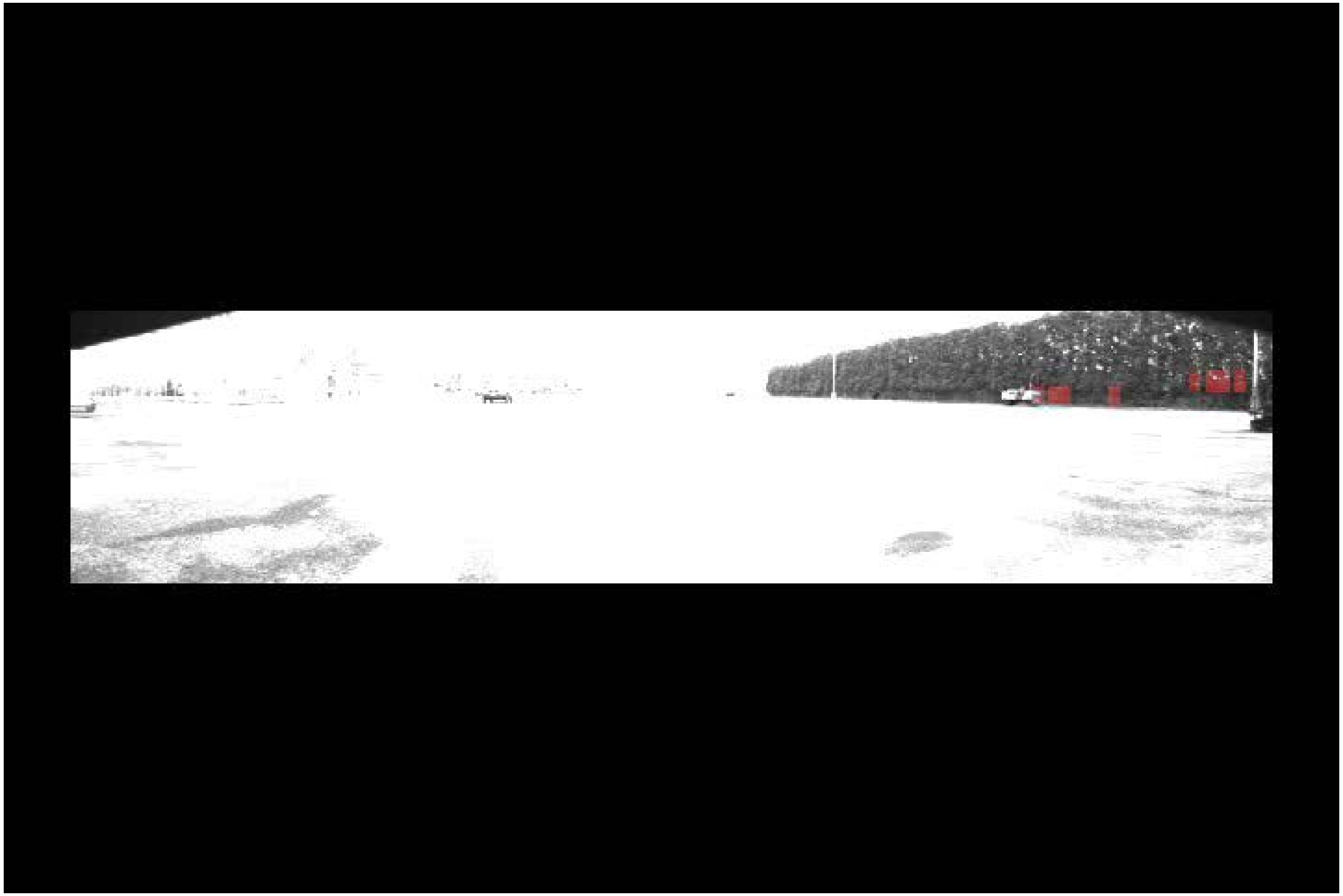
80 Deg

Vehicle Classifier

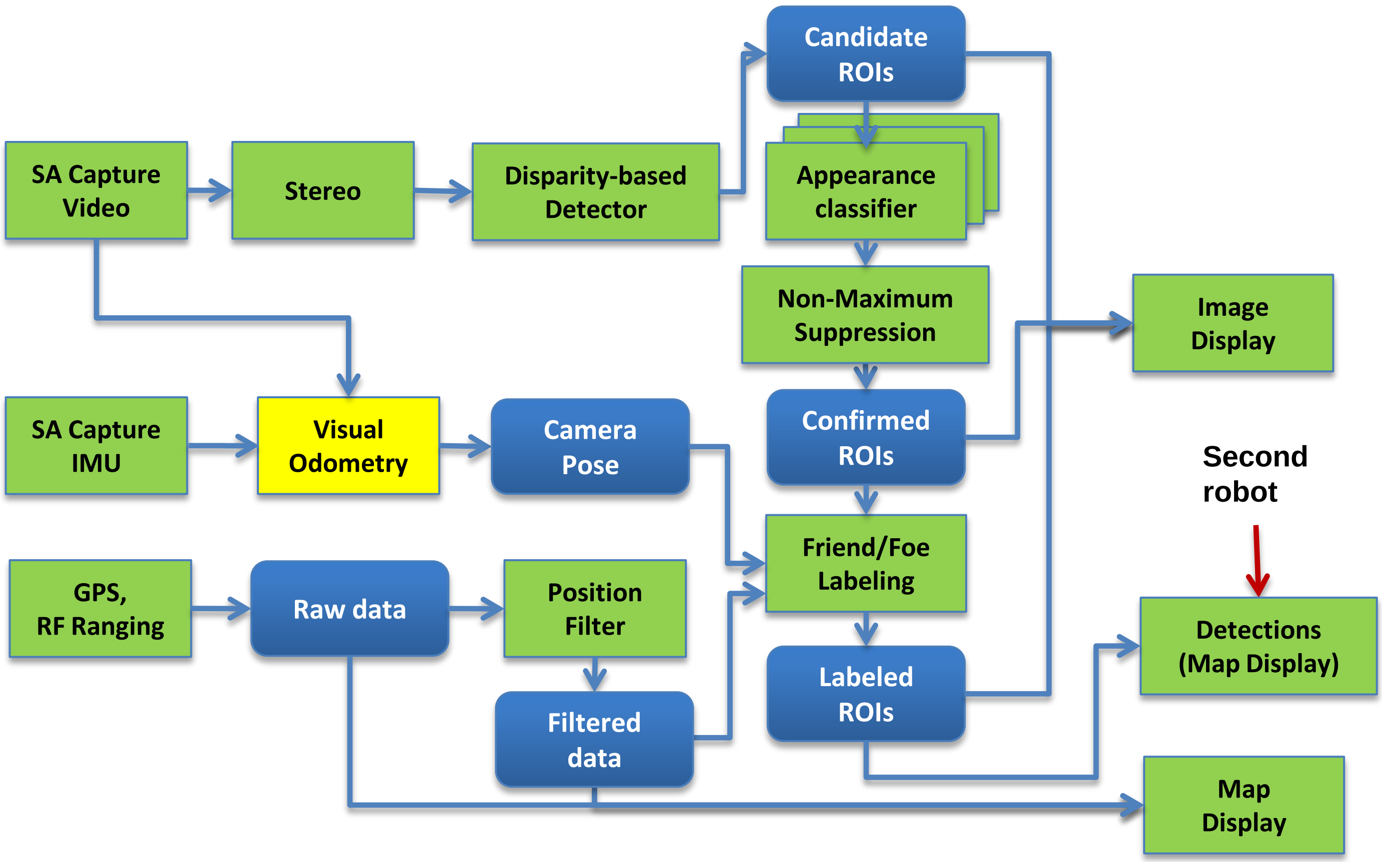
- **Vehicle appearance** varies widely due to viewpoint, body type, occlusion.
- **Our solution**: A second Convolutional Neural Network (ConvNet) was trained to recognize vehicles.
  - Can learn extreme variability in object appearance
  - Fast runtime performance
  - Trained on raw data without extensive preprocessing or parameter tuning
- **Network details**: The vehicle ConvNet is similar to the pedestrian ConvNet:
  - 6 layer hierarchy (3 convolutional layers, 2 pooling layers, and a fully connected layer)
  - 60x30 pixel field of view
  - 12,000 trainable parameters.
- **Training process**: Based on human-annotated videos
  - 580,000 labeled **positives** (ROIs with vehicles) and **negatives** (ROIs with no vehicles)
  - Network parameters are optimized using stochastic gradient descent

Vehicle Classifier – Qualitative Results

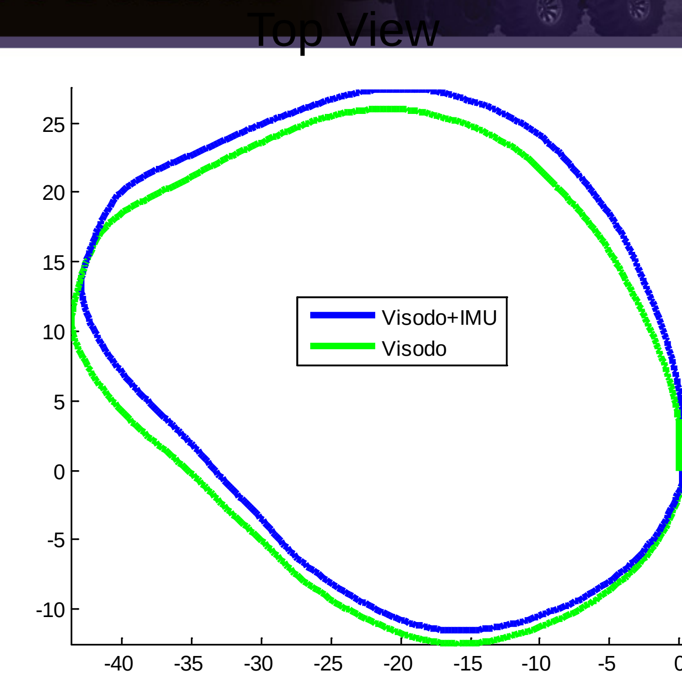


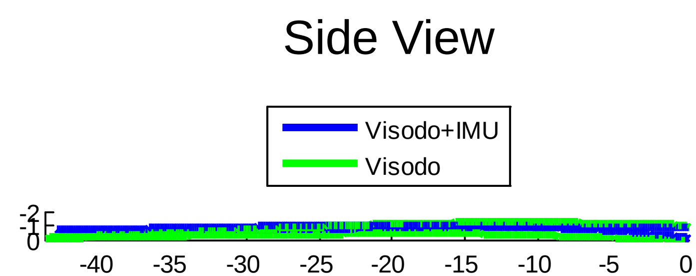
Visual Odometry

ROBOTIC SYSTEMS



Outdoor Loop Closure Test



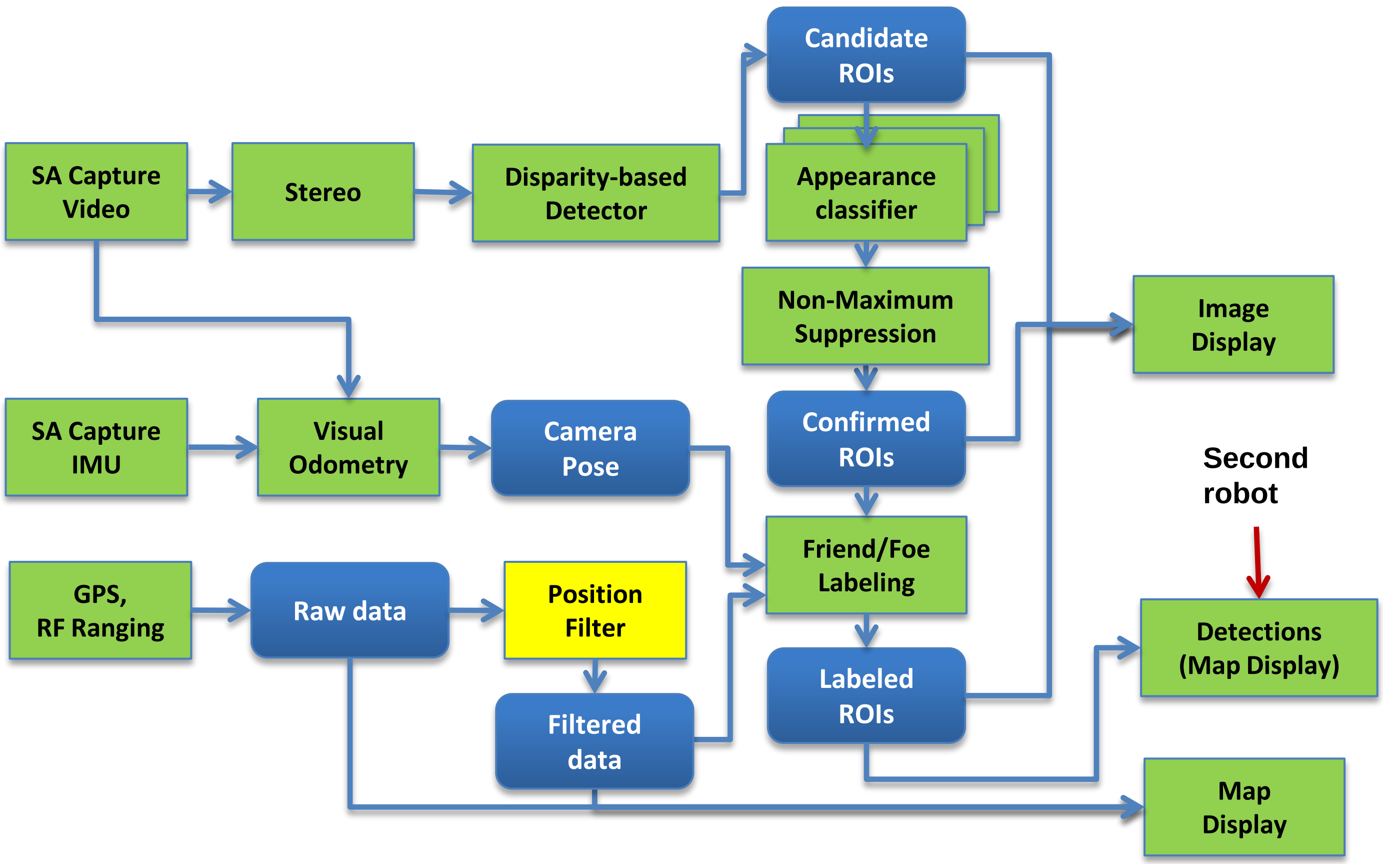


FishEye Loop Closure Test Results

Outdoor	Total Travelled Distance (meter)	Loop Closure Error (meter)	Drift Rate (%)
Loop 1 Visodo	124.9396	1.1138	0.89
Loop 1 Visodo+IMU	124.0460	1.0812	0.87
Loop 2 Visodo	122.4757	0.8724	0.71
Loop 2 Visodo+IMU	122.3237	0.7168	0.58

Indoor	Total Travelled Distance (meter)	Loop Closure Error (meter)	Drift Rate (%)
Loop 1 Visodo	51.2833	0.4648	0.91
Loop 1 Visodo+IMU	51.3082	0.3699	0.72
Loop 2 Visodo	105.9501	0.5210	0.49
Loop 2 Visodo+IMU	105.9180	0.5015	0.47

RF+GPS Position Filter



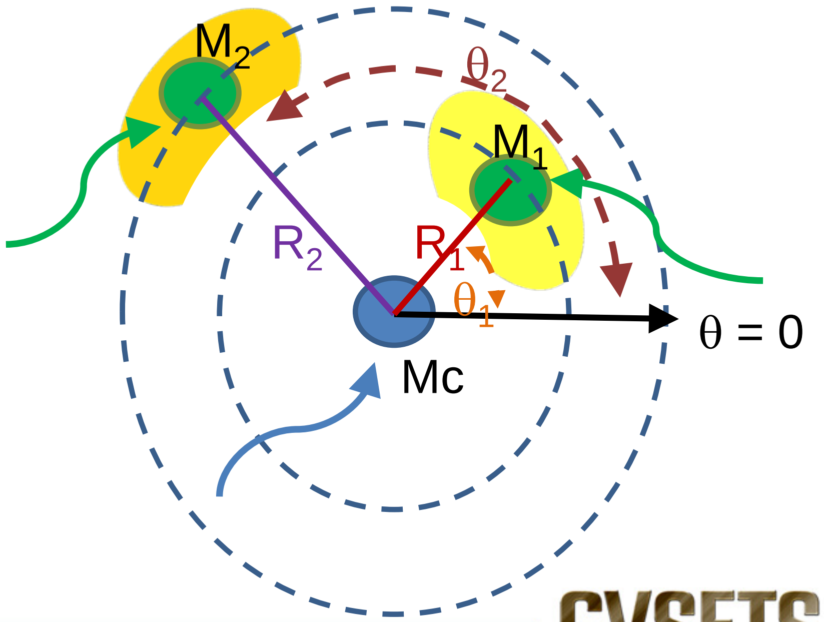
Notation

$(M_c, M_1, M_2, M_3, \ldots M_m)$: m+1 mobile nodes, M_c is the central Visodo/IMU/GPS/RF node. Other nodes are GPS/RF nodes.

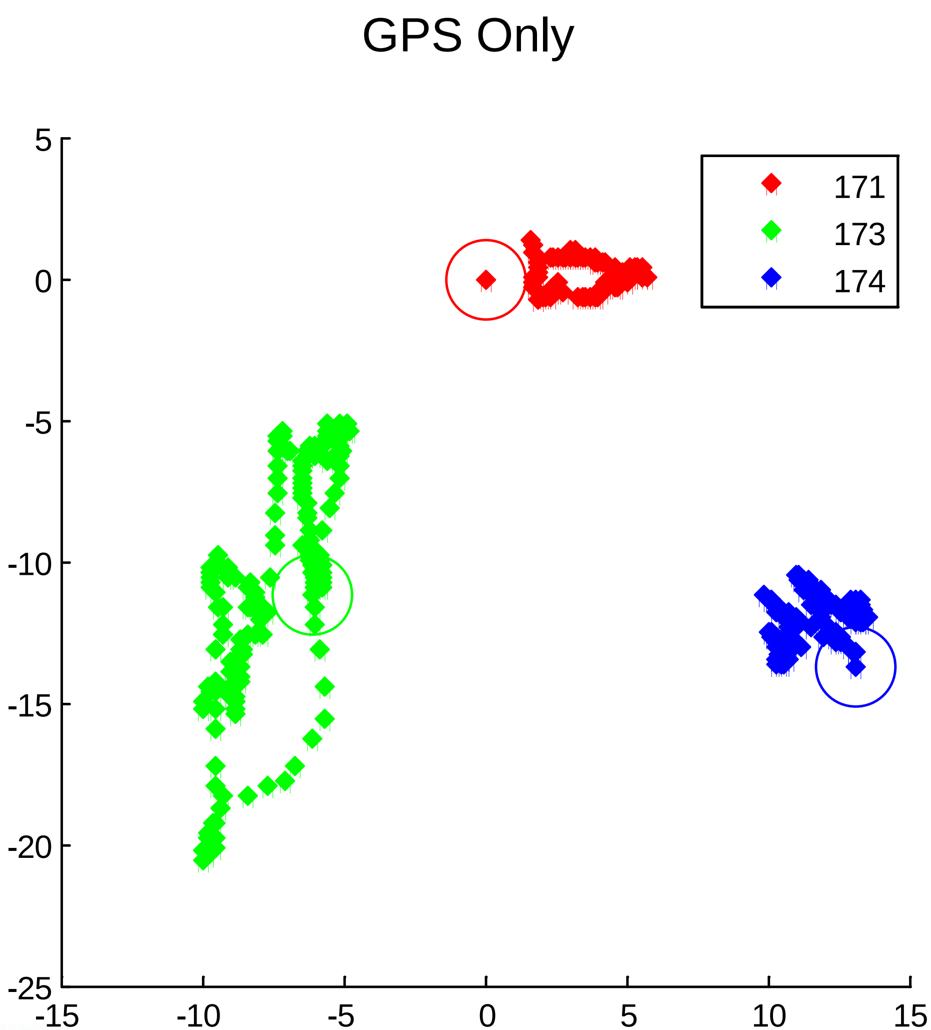
M_c - (X_c, Y_c, V_c^X, V_c^Y) : The simplified representation from our error-state EKF
M_i - $(R_i, \theta_i, V_i^X, V_i^Y, b_i)$: A normal EKF (no IMU, odometry) but in “relative-polar”(RP) coordinate system. The origin is the position of M_c, which can move.

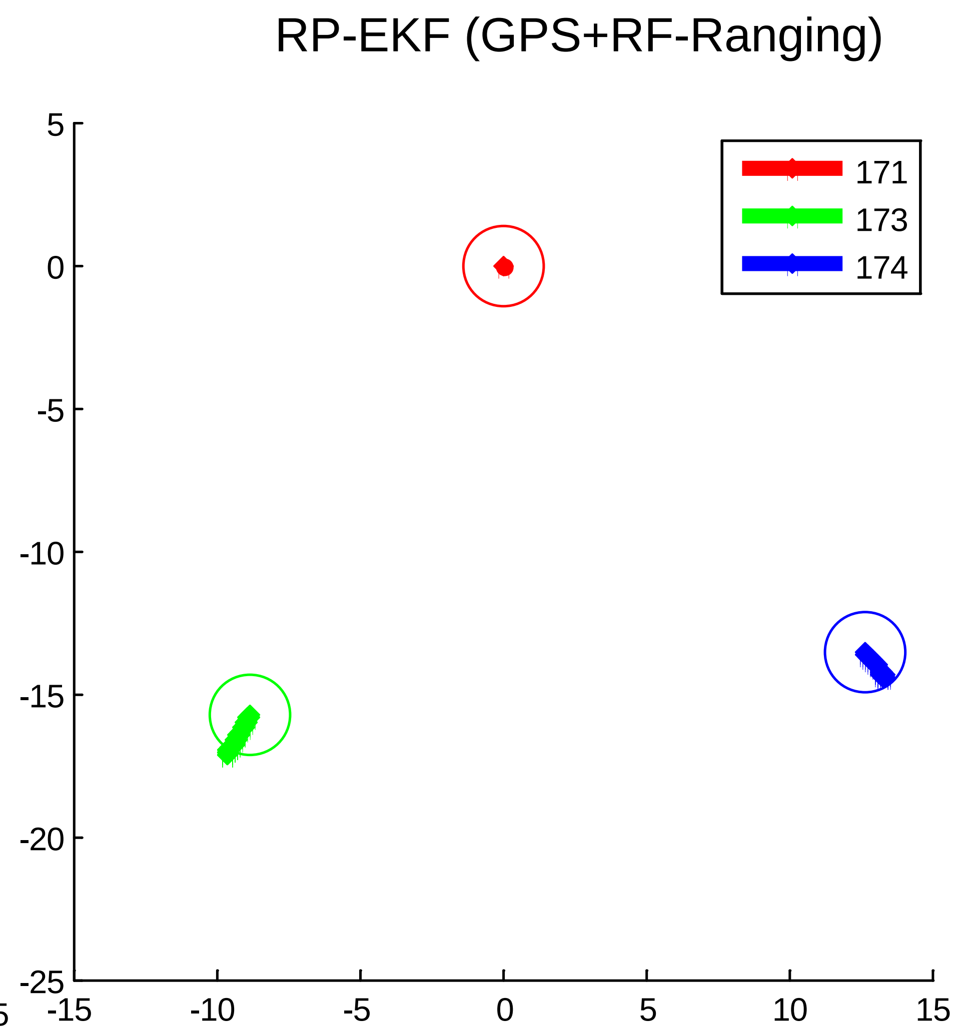
Polar representation is less used in EKF, but recently has been proved to be better suited for applications such as navigation with mapping of static RF-ranging nodes.

We developed a new relative-polar formulation in EKF for our application (moving RF-ranging nodes, no odometry information).



Three Static Nodes - 2011.01.20-14.24.27





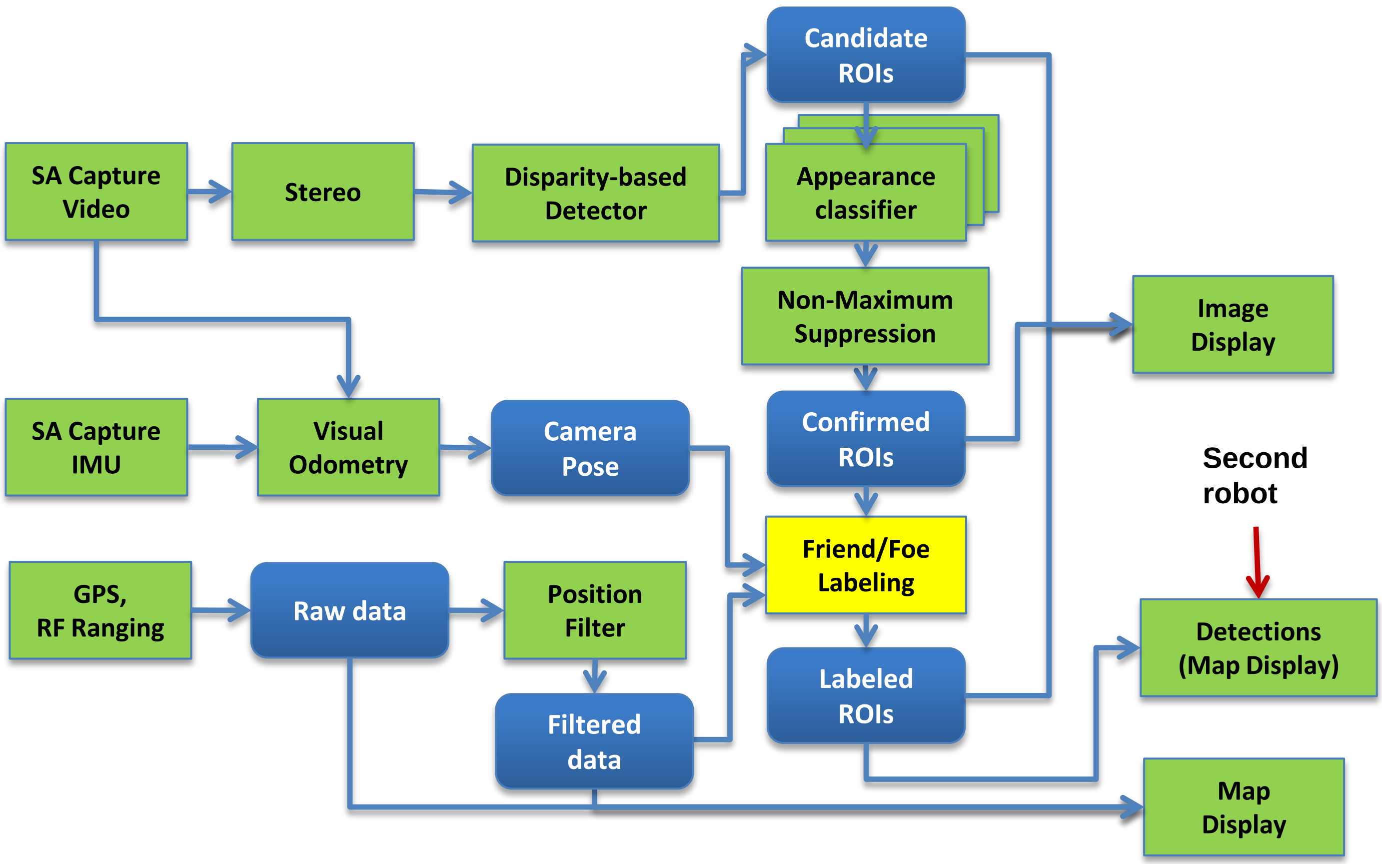
Loop closure test

	Loop Closure Error (meter)
Visodo	0.2554
GPS	7.6335
GPS+RF	4.3725

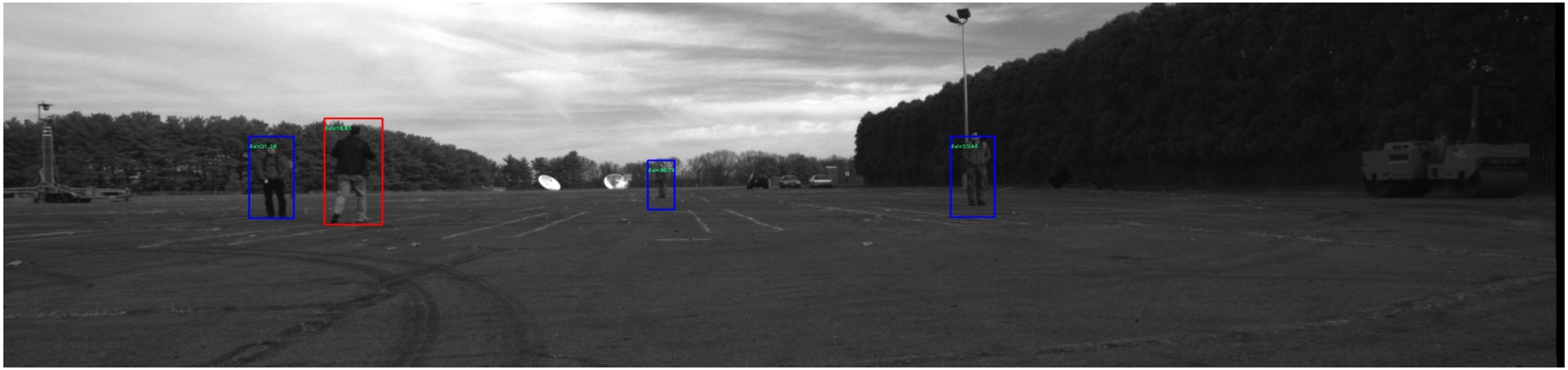
Travelled Distance:
74.69 meters

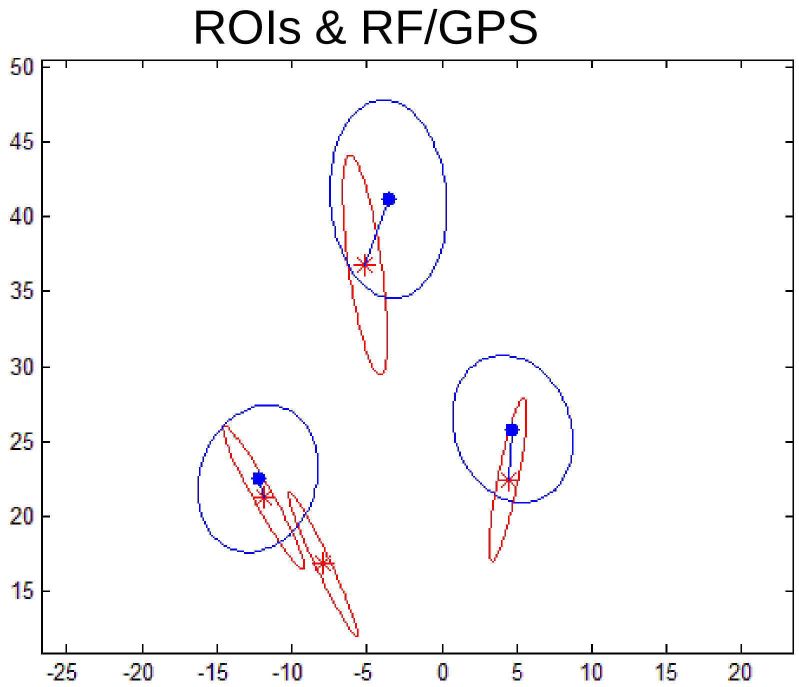
Blue: Visual Odometry
Yellow: GPS
Green: RF + GPS

Friend/Foe Labeling



Association Example





* People ROIs
• RF/GPS

Image and map displays

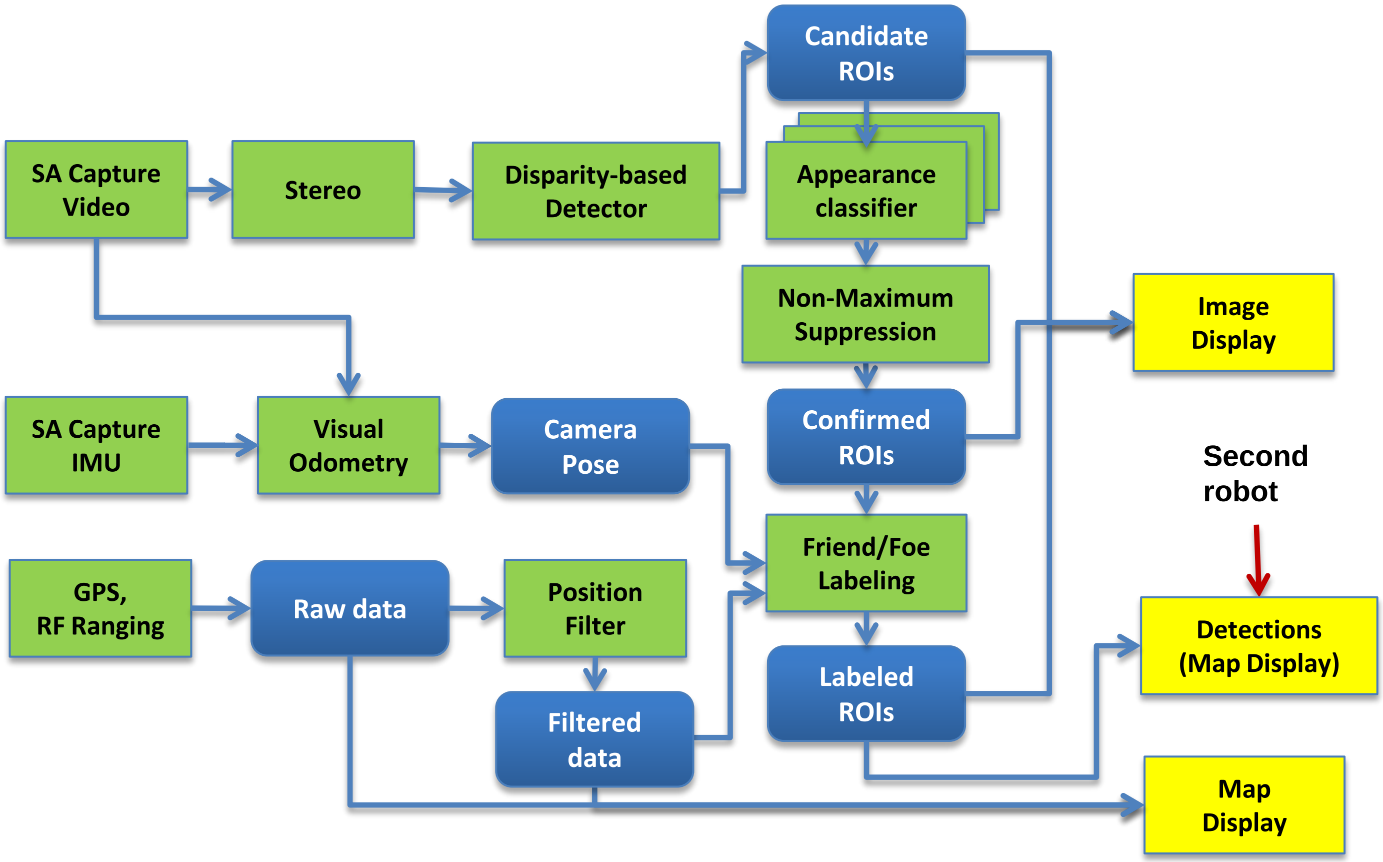
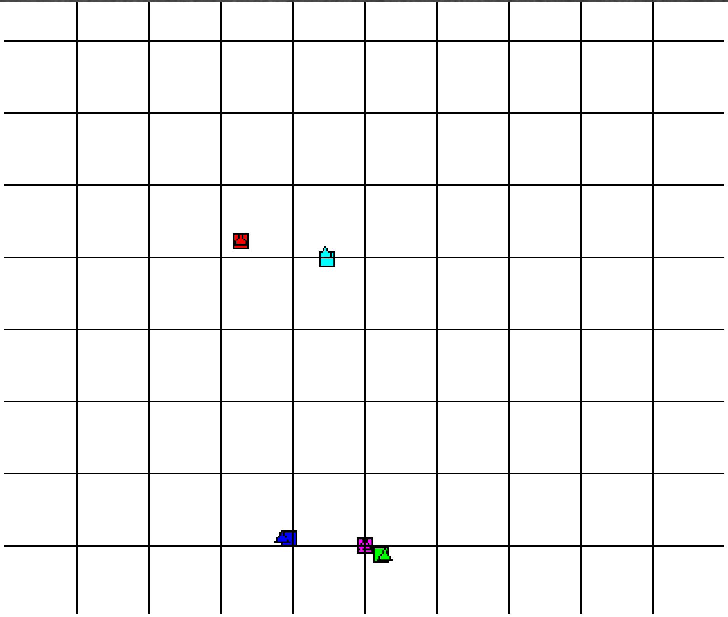
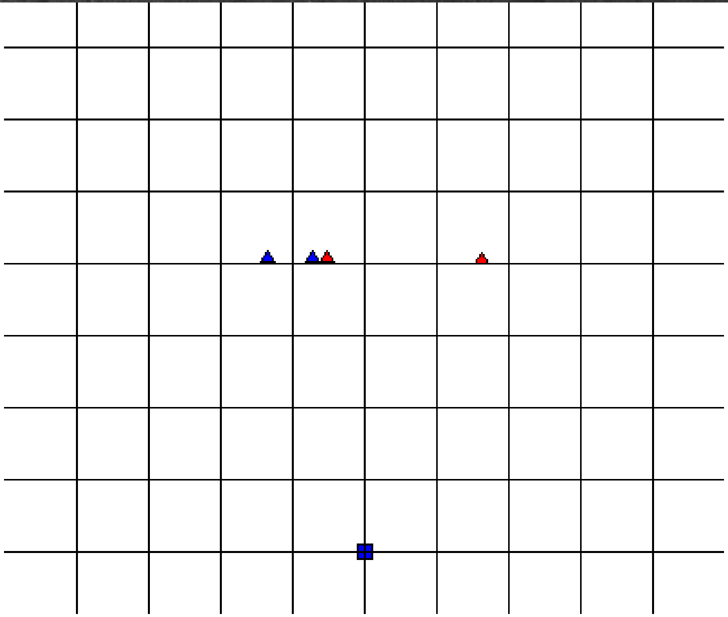


Image display

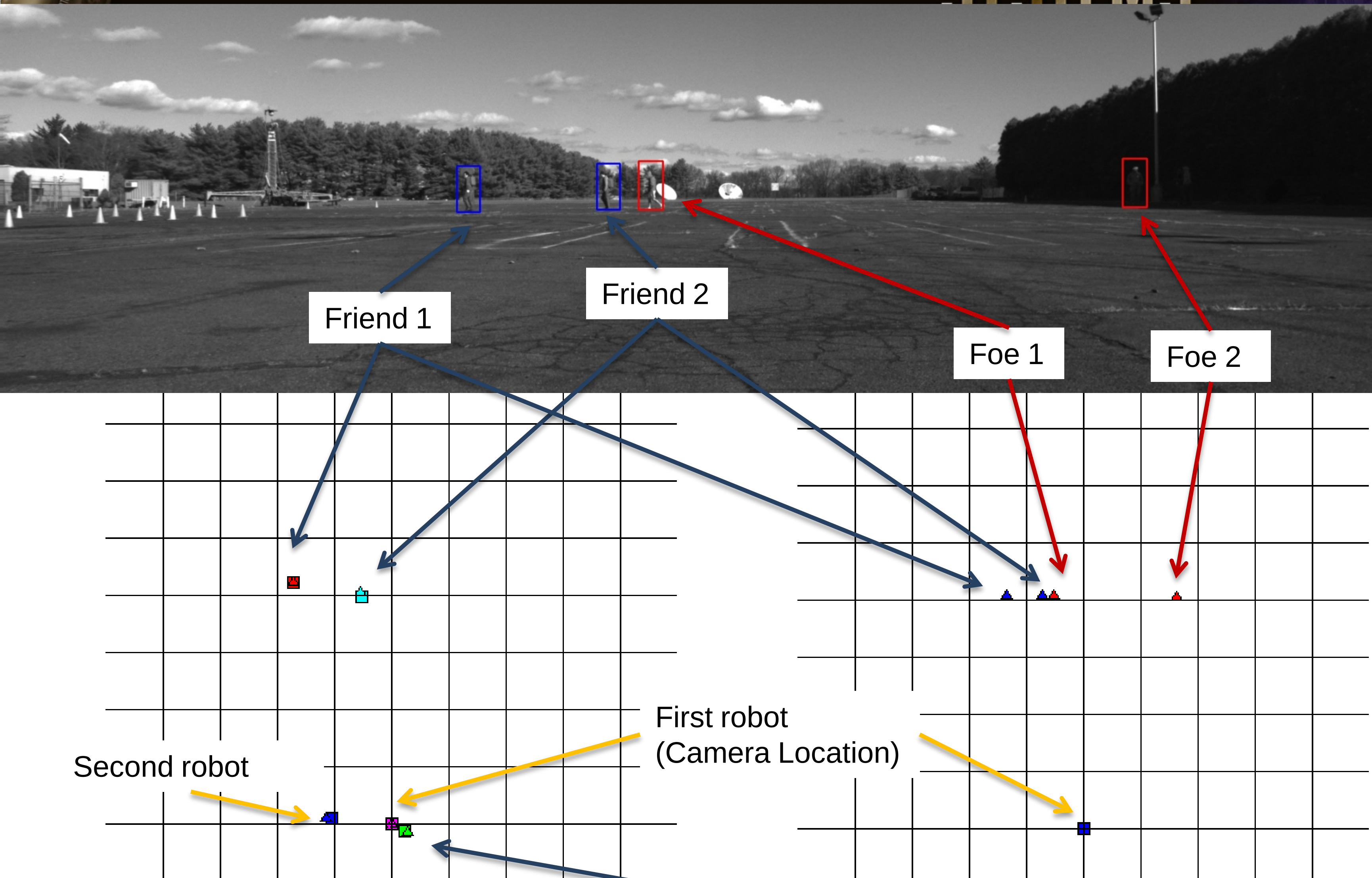


Map display



Detection display

Image display



Friend 3 (outside camera Field of View)

Map display

Detection display

Baseline Testing

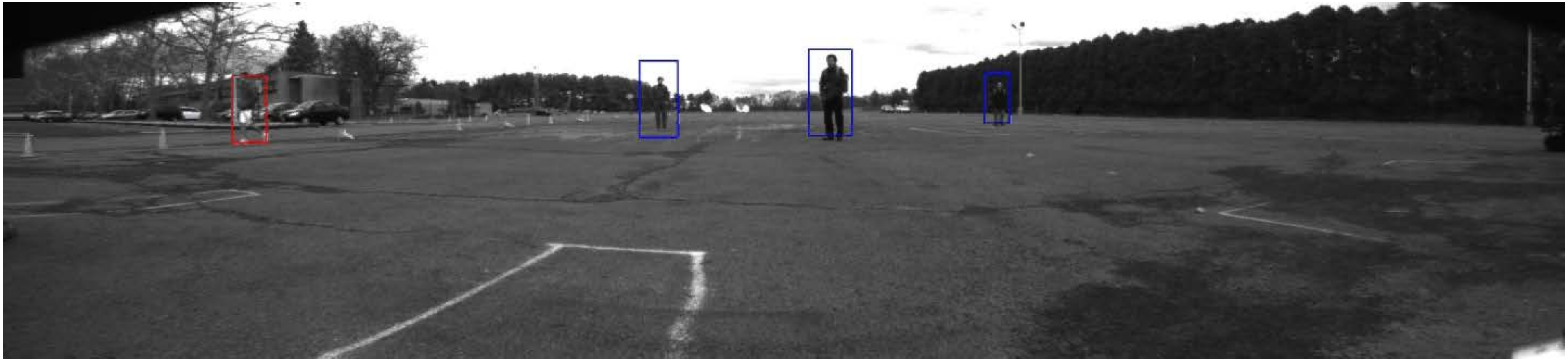
- Baseline testing for the system was performed with combination of Friends, Foes and vehicles at varying distances.
- The Friends (up to three) and Foes (up to six) were systematically tested in varying combinations moving in front of the robots at ranges from 10 to 100 meters.
- The Friends/Foes varied in speed and motion from a slow crawl to a fast sprint.
- Similar testing was then preformed with automobiles. One to three vehicles varying from parked to moving at 25 mph at ranges from 10 to 100 meters.
- The EETs then became more complicated. Introducing various sets of Friends, Foes and vehicles in random patterns to try and find the failure point of the system.

Baseline Testing

2 Friends, 1 Foe at 40m (80deg camera)

Multiple tests (Fisheye camera)

ROBOTIC SYSTEMS

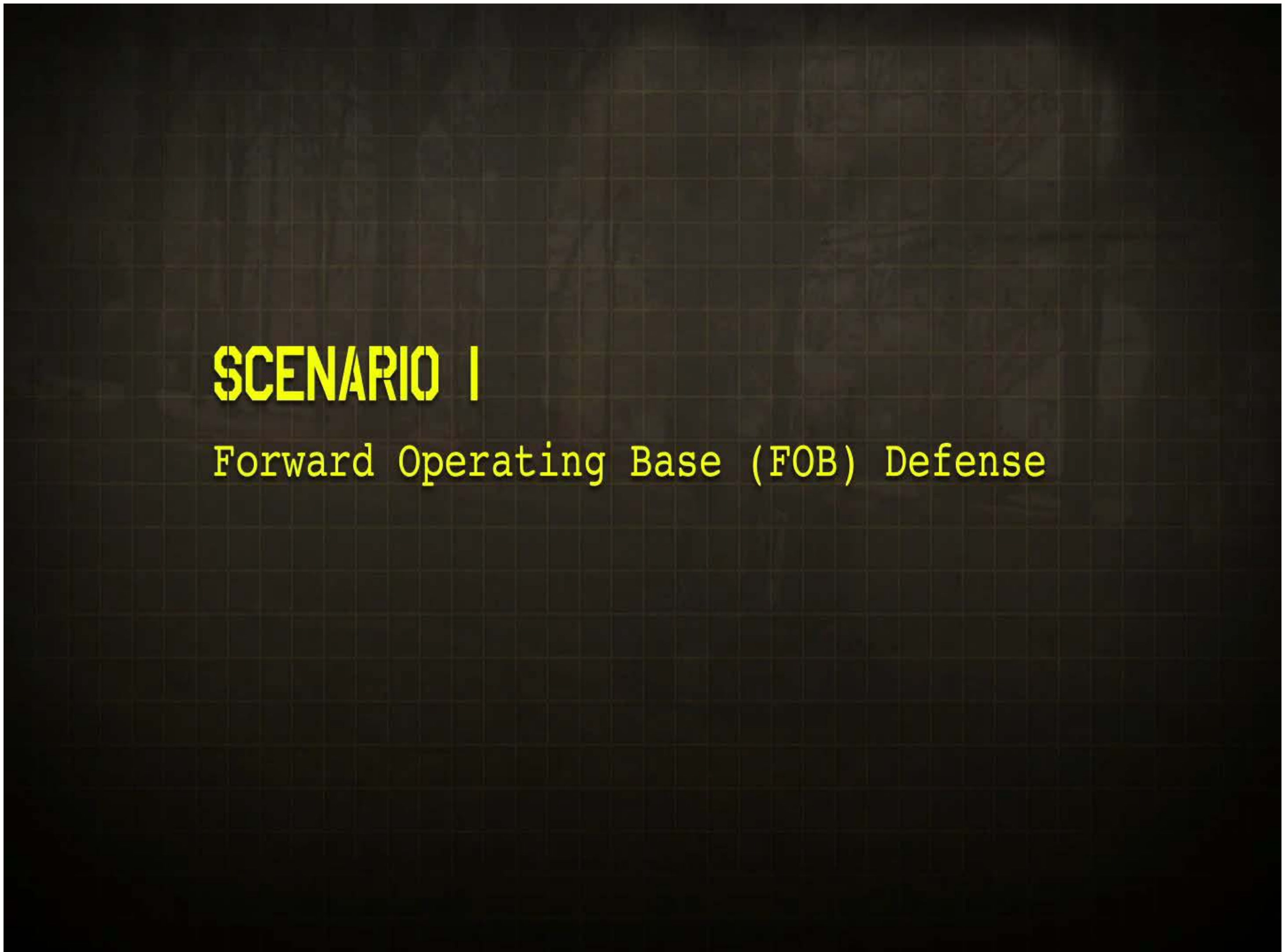


- Several tests:
  - Three friends at ~20m
  - Foes at 10, 20, 30 and 40m
  - Friends at 20 and 50m

Scenario Testing

- The first Scenario was setup with friendly forces being dug into their fighting positions with two fixed Combat Identification robots monitoring the fields of fire. The enemy could attack at any moment and the robots would have to identify if the personnel approaching the FOB were friendly forces or enemy forces before any friendly forces could engage the target.
- The second scenario was identical to the first scenario with the exception that one of the Combat Identification robots could move across the field of fires in order to establish a better line of sight to identify the targets as friendly or enemy threats.
- In the third scenario, the friendly forces conducted patrols from the FOB to a local village; upon returning from the mission the two fixed Combat Identification robots would have to identify the objects as friendly before access would be allowed into the FOB.

Scenario Testing



Conclusions

- There is a need to increase available resources by eliminating tasks that are conducted by humans and having robots complete these tasks. The Combat ID system addresses this need by allowing for a broader field of view/line of sight and object movement detection then one single person can accomplish.
- The CombatID program successfully showed that a unmanned robotic equipped with the CombatID payload could scan the same line of sight as a Solider.
- As Soldiers and commanders become more accustomed robots on the battlefield, the acceptance and utility of CombatID like capabilities will become combat multipliers for the operational commander.